

# Computational Storage and AI

Run your AI models directly on your disk

## Our latest CSD

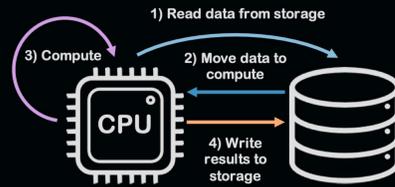
- 256 GB Storage – Extensible
- PCIe 3.0 x4 – NVMe standard
- Quad-core Cortex-A76 + Quad-core Cortex-A55 ARM Processors
- 16 GB DDR4 RAM
- 6 TOPS Neural Processing Unit
- 610 GFLOPS Graphics Processing Unit
- USB Type C and PCIe for extensions
- Linux 6.7 kernel
- Network over NVMe, no extra cable
- Can run full OS (e.g., Ubuntu/Debian)
- Low-power < 16W
- Low-cost < 200\$



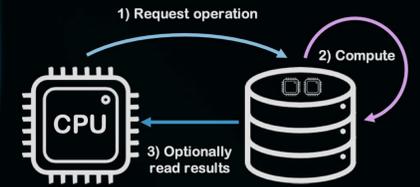
## 1 Reduce data movement

Computational Storage Devices (CSDs) seamlessly integrate processing and storage in a single compact unit. By enabling on-the-spot computations at the data source, CSDs eliminate bottlenecks between processors and storage, optimizing performance and reducing the global energy footprint.

### Traditional architecture



### Computational Storage



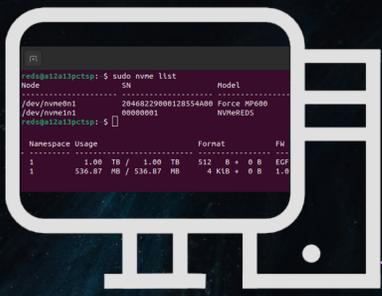
## 2 Scale with your data

Escape the limitations of traditional processors with CSDs, where processing power scales alongside storage capacity. CSDs can also interact with other traditional disks and even extend storage. Moreover, CSDs can intelligently perform computations even during the main processor's downtime, making them ideal for lengthy tasks like model training and video transcoding.

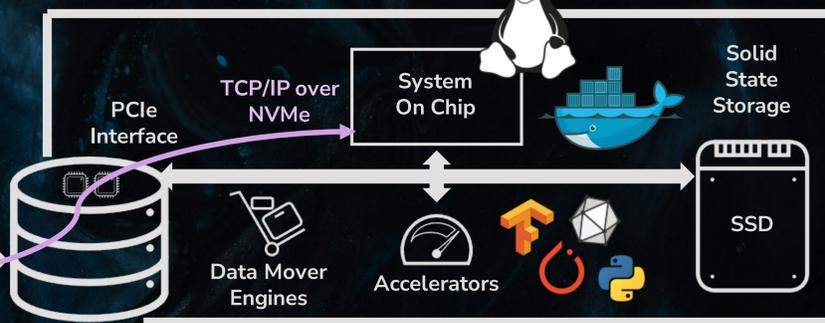
## 3 Standard environments & open-source



### Normal PC

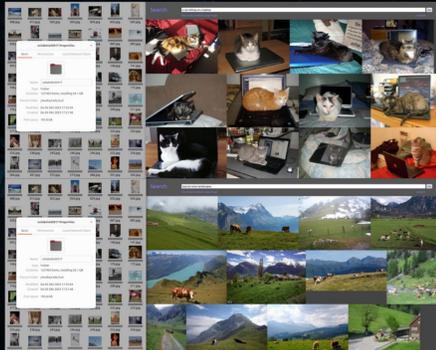


### Anatomy of a CSD



Our CSDs are built around an embedded Linux kernel and offer a rich environment to offload computational functions. The PC sees the CSD as an NVMe drive, TCP/IP is tunneled over NVMe to enable communication without the need for any extra cables.

### Computational Storage Devices & Storage Extension (Samsung SSD for scale)



By combining natural language processing (NLP) and image recognition we can make our computational storage drive content aware!  
We can query the disk to find "cats sitting on laptops" or "typical Swiss landscapes" in very large image databases

## 4 Ready for big-data AI, distributed ML, and more !

Our CSDs allow to run and train AI models at the data source. We support GPU, FPGA and Neural Processing Unit hardware acceleration. Explore novel ways to interact with your data; For instance, use natural language processing and image recognition to query extensive image databases in English, all processing done directly on disk !

Rick Wertenbroek, Peter Podolec, Roberto Rigamonti, Enrico Petraglio, Yann Thoma, & Alberto Dassatti



Let's talk!



REDS  
Institut  
Reconfigurable  
and Embedded  
Digital Systems